

4G-5

SL-TRANS における文節音声認識

- HMM 音韻認識と LR 構文解析法による文節音声認識 -

北 研二 坂野 俊哉 保坂 順子 川端 豪

ATR 自動翻訳電話研究所

1 はじめに

本稿では、音声言語日英翻訳実験システム SL-TRANS の音声認識部について説明する。音声認識部は、文節単位に発声された日本語音声を認識し、複数の認識候補(文節ラティス)をスコア付きで出力する。

音声認識の方式として、ATR 自動翻訳電話研究所の提案した HMM-LR 法を用いている。HMM-LR 法は、言語の構文情報を用いて音声データを直接解析する方式であり、この方式を用いることに

より高精度のかつ効率的な認識処理系を構成することができる。

また、HMM-LR 法の出力である各認識候補に付属している認識スコアの統計特性値を用いて、認識候補の絞り込みを行っている。

以下で、HMM-LR 法および統計処理による認識候補の絞り込みについて述べる。

2 HMM-LR 法^[1]

HMM-LR 法は、言語の構文情報を用いて音声データを直接解析する方式である。構文情報の取り扱いには拡張 LR 構文解析法^[2]を用いており、

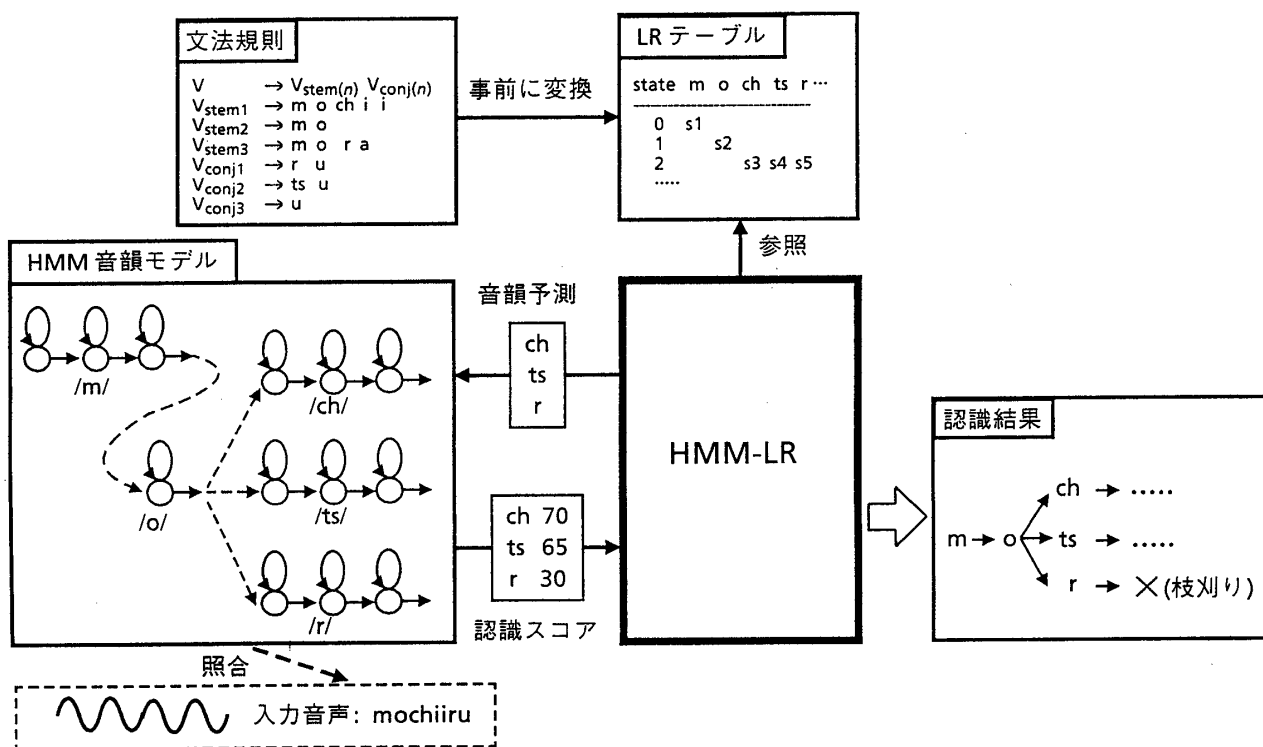


図1 HMM-LR 音声認識システム

Speech Recognition Method in SL-TRANS

Kenji Kita, Toshiya Sakano, Junko Hosaka, Takeshi Kawabata

ATR Interpreting Telephony Research Laboratories

音韻認識部に HMM (Hidden Markov Model) [3] を用いている。

HMM 法は、近年注目をあびている音声認識の手法である。HMM は音韻のゆらぎを統計的に表現できるという特徴があり、発声状況やコンテキスト等の違いによる音韻変動に対して robust なモデルを構成することができる。また、認識単位を音韻に設定しておけばこれを基に任意の単語モデルを合成することができるため、大語彙の音声認識に適している。

図1に、HMM-LR法に基づく音声認識システムの構成図を示す。HMM-LR 音声認識システムは、拡張 LR 構文解析法で用いられる構文解析動作表から入力された音声データ中の音韻を予測し、予測された音韻の尤度を HMM 音韻照合で調べることにより、認識を進めていく。いま、図1において /mo/ という音韻列が既に認識されているとする。このとき、システムは構文解析動作表の状態2にいることになり、状態2で shift 動作の指定されている音韻 /ch/, /ts/, /r/ を予測し、これらの予測された音韻の HMM モデルを用いて各音韻に対する尤度を調べる。このように、構文解析動作表は音韻予測に用いられるため、文法規則の終端記号は、単語や品詞名ではなく、音韻となっている。

一般に、構文解析動作表から予測される音韻は複数個あるので、認識途中では多数の認識候補が残される。探索空間を縮小するために、各音韻列に付与されている尤度を用いて、ビームサーチ法による枝刈りを適宜行っている。

HMM-LR 音声認識システムは、現時点では特定話者を対象としている。システムの認識性能は、約 1000 語を対象とした文節認識実験で、文節認識率 87.1%、上位候補 5 個までの累積認識率 97.7% を達成している(セパレートベクトル量子化による HMM を用いた場合の3名の話者の平均値)。認識誤りの傾向として、付属語部分の脱落および挿入誤り、発音の似た自立語間での置換誤りが比較的多くみられる。文節の認識例を図2に示す。

正解:	会議に		参加したいのですが
1:	会議に	(0.94)	参加したいのですが (0.84)
2:	会議	(0.06)	参加したいのですか (0.13)
3:			参加されますか (0.03)

図2 文節の認識例

3 統計処理による絞り込み [4]

HMM-LR 音声認識システムは、一般に複数の認識候補を出力するが、各候補に付属している認識スコアの統計特性値により、認識スコアの信頼性評価を行い、候補の絞り込みを最終段階で行っている。

実際には、各文節ラティスごとにラティス中の認識スコアの分散値および最大2階差分値を用いて、完全認識ランク(正しい候補を包含する最低ランク)を決定している。例えば、分散を用いた場合には、ラティス中の認識スコアの分散値 v に対する完全認識ランク $rank(v)$ は、 d を分布パラメータとするとき、次式で決定される。

$$rank(v) = -5v/d + 5$$

統計処理による絞り込みにより、文節ラティス中の認識候補を平均 2.1 個に絞ることができた。また、この方法はラティス中に正解がない場合を検出することにも適用できる。この場合、SL-TRANS ではシステム側から人間に問い合わせを行う。

4 おわりに

音声言語日英翻訳実験システム SL-TRANS の音声認識部について述べた。音声認識の方式には、文節内の構文情報を用いた HMM-LR 法を採用しており、認識スコアの統計特性値を用いた候補の絞り込みをあわせて行っている。

謝辞

研究の機会を与えて頂いた ATR 自動翻訳電話研究所 樽松明社長に深謝致します。また熱心に討論して頂いた同研究所 音声情報処理研究室 およびデータ処理研究室の皆様へ感謝致します。

参考文献

- [1] 北,川端,斎藤:「HMM 音韻認識と予測 LR パーザを用いた文節認識」,電子情報通信学会技術研究報告 SP88-88 (1988).
- [2] Tomita,M.: *Efficient Parsing for Natural Language: A Fast Algorithm for Practical Systems*, p.201, Kluwer Academic Publishers (1986).
- [3] 中川:「確率モデルによる音声認識」,電子情報通信学会 (1988).
- [4] 坂野,北,森元:「音声認識候補の統計処理による絞り込み」,電子情報通信学会春季全国大会 A-21 (1989).